Sugden, P., & Canagarajah, CN. (2003). Blind signal separation using fixed overcomplete basis function dictionaries. In *2003 International Symposium on Circuits and Systems, 2003 (ISCAS '03)* (Vol. III, pp. III-44 - III-47). Institute of Electrical and Electronics Engineers (IEEE). https://doi.org/10.1109/ISCAS.2003.1204951

Peer reviewed version

Link to published version (if available):
10.1109/ISCAS.2003.1204951

BLIND SIGNAL SEPARATION USING FIXED OVERCOMPLETE BASIS FUNCTION DICTIONARIES

Paul Sugden and Nishan Canagarajah

Digital Music Research Group
Merchant Venturer's Building
Woodland Road
Bristol, BS8 2BL

ABSTRACT

A solution for achieving blind separation for underdetermined systems is to use an overcomplete basis function set that has the ability to span all possible inputs. Ideally, such a basis would be learned for each set of inputs but this is computationally expensive. A less processor intensive system is shown using a fixed dictionary of basis functions learned from existing sources and reduced using a correlation-based method. The relation between dictionary size and separation performance for underdetermined scenarios is examined and we demonstrate that a reduced dictionary can produce comparable results using less computational power.

1. INTRODUCTION

The use of independent component analysis (ICA) ([1],[2]) for blind separation has been well documented. However, the use of traditional ICA is restricted to systems where there are at least as many sources as sensors, i.e. a determined system. For the underdetermined case, a more general approach must be found. An idea of this nature has been presented in ([3]) by learning an overcomplete basis set from the system outputs and has been used for blind separation ([4]). However, the learning process is complex and may be unsuitable for real-time use. Here, we will describe the use of an overcomplete solution using basis functions learned using the aforementioned method from existing speech samples and then reduced by eliminating highly correlated elements. This reduced basis set can then be used for blind separation.

A similar approach is used in vector quantization to build libraries for signal compression ([5]).

The data model for blind separation we define is of a system where mixing of signals is linear and stationary. From N sources $s_{1\ldots n}(t)$ we get the M outputs $x_{1\ldots m}(t)$. This gives equation 1.

$$\mathbf{x}(t) = \mathbf{A} \cdot \mathbf{s}(t) \tag{1}$$

where $\mathbf{A}$ is an M by N mixing matrix. We therefore need to find the inverse of matrix $\mathbf{A}$, For a determined system, $M \geq N$ and $\mathbf{A}$ has full rank. The recovered sources $y_n(t)$ can then be obtained from $\mathbf{y} = \mathbf{W}\mathbf{x}$. For classical ICA, the condition of $M \leq N$ must be satisfied; there must be at least as many outputs as inputs.

2. LEARNING AN OVERCOMPLETE BASIS

In order to solve the underdetermined case we use an overcomplete set of basis functions similar to to the Gabor set [5]. However, whilst this set is mathematically derived, our basis function set is learned using the algorithm described in [3]. The task is to find an optimum basis function set for a given signal based on the data model shown in equation 1.

It is assumed the sources s_i are mutually independent such that the joint probability distribution has the form $P(\mathbf{s}) = \prod_{i=1}^{M} P(s_i)$ and each source s_i has a sparse distribution (i.e. super-Gaussian). A probabilistic approach to inferring the sources is based on the maximum a posteriori value of s:

$$\hat{\mathbf{s}} = \max_{\mathbf{s}} P(\mathbf{s}|\mathbf{x}, \mathbf{A}) = \max_{\mathbf{s}} P(\mathbf{x}|\mathbf{A}, \mathbf{s})P(\mathbf{s}) \tag{2}$$

Given basis vectors $\mathbf{A}$ and observations $\mathbf{x}$, equation 2 can be optimized using linear programming methods (see [3] for more detail). The objective for learning the basis vectors $\mathbf{A}$ is to maximize the probability of the data:

$$P(\mathbf{x}_1 \ldots \mathbf{x}_T|\mathbf{A}) = \prod_{i=1}^{T} P(x_i|\mathbf{A}) \tag{3}$$

Computation of likelihood requires marginalizing over all possible sources as in (4)

$$P(\mathbf{x}|\mathbf{A}) = \int P(\mathbf{x}|\mathbf{A}, \mathbf{s})P(\mathbf{s})d\mathbf{s} \tag{4}$$

For the special case of zero noise and $\mathbf{A}$ being invertible (a complete basis), this equation is solvable and leads to the

0-7803-7761-3/03/$17.00 ©2003 IEEE

standard independent component analysis (ICA) algorithm. In [3], equation 4 is approximated by fitting a multivariate Gaussian around $\hat{\mathbf{s}}$. The learning rule is:

$$\Delta\mathbf{A} = \mathbf{A}\mathbf{A}^T \frac{\delta}{\delta\mathbf{A}} \log P(\mathbf{x}|\mathbf{A}) \approx -\mathbf{A}(\phi\hat{\mathbf{s}}\hat{\mathbf{s}}^T + \mathbf{I}) \quad (5)$$

where $\phi(\hat{s}_i) = \delta \log P(s_k)/\delta\hat{s}_i$ and is the cost function. The matrix $\mathbf{A}$ is not restricted to be square and thus works as an overcomplete representation.

3. BUILDING THE SIGNAL DICTIONARY

To build the signal library, a selection of p speech signals from the same speaker were analyzed using the algorithm described in [3] and a basis function set of q was obtained from each. These sets were combined to produce a set of $z = p \times q$ functions $\mathbf{\Phi} = (\phi_1 \ldots \phi_z)$, where ϕ_i represents the i-th basis function. A correlation table was built, whereby individual functions were compared with each other function. Each of the functions was then rated by the highest l correlations found with another function as shown in (6)

$$\Lambda = \max_{1\ldots l}(\mathrm{corr}_{i,j}(\mathbf{\Phi}^T, \mathbf{1}\cdot\mathbf{\Phi})|_{i\neq j}) \quad (6)$$

The array $\mathbf{1}$ represents a square matrix of 1s of size z by z and $i = 1\ldots z$ and $j = 1\ldots z$. In the array Λ, two types of values are stored; the absolute correlation value and the value of j where this value is found, thus creating a reference between correlated elements ϕ_i and ϕ_j. Its structure can therefore be considered $\Lambda = (\mathrm{correlations}_{1:l}|\mathrm{positions}_{1:l})$.

The complete dictionary is used to decompose a previously unused speech extract by the same speaker. The error at the end of decomposition is recorded and the size of the library decremented by one. The function removed is decided by the following process; first the highest correlation value is found in the first column of Λ (7) and the row where this maximum occurred stored in k. We must consider that the highest correlation value may be held by two functions as a result of a correlation with each other. In this case, the values of the second highest correlations (in the second column of Λ) are compared and the higher of this comparison removed.

$$k = \max_{i,1} \Lambda \quad (7)$$

If the correlation value from the corresponding element found in $r = \Lambda_{k,1+l}$ had not been eliminated it must also be removed. The elements in row r are shifted left in order to position the next highest correlation value in the first column. (8).

$$\Lambda_{r,i} = \Lambda_{r,i+1}|_{i=1\ldots l} \quad (8)$$

Finally, to remove the basis function from further analysis, the coefficients of row k are set to zero as shown in (9).

$$\Lambda_{k,1\ldots 2l} = 0 \quad (9)$$

The results of this procedure were plotted as the number of basis function against the decomposition error. An example of such a plot can be seen in Figure 3. Using the graph, we are looking for a point where the trade-off between error and the number of dictionary elements.

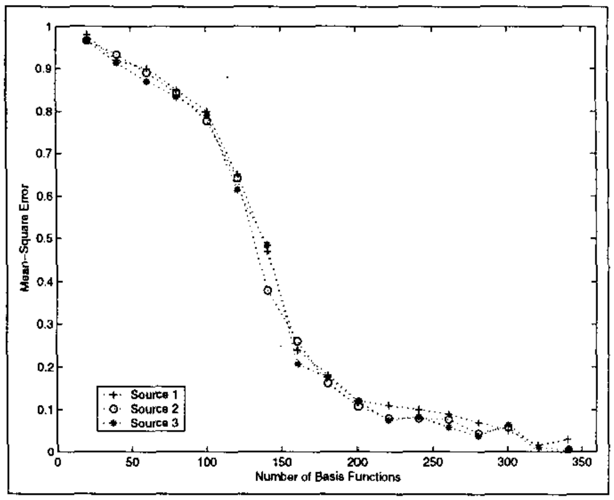

Figure 1: Plot showing error as a function of the number of basis functions used in the decomposition of a speech sample using a learned library created from the same speaker.

The example graph shows an appropriate 'knee' in the error at around 150 basis functions for all three sources and this is therefore the size of the dictionary we choose. The blind separation is performed by formulating the decomposition of the underdetermined mixtures as a linear programme that searches for the *sparsest* possible representation of the data ([6]), i.e. one that uses the least number of basis functions.

4. PERFORMING BLIND SEPARATION

In the case of finite noise, ξ, there is a general method of optimizing the s, using the gradient of the log posterior in an optimization algorithm. An alternative can be used when the prior is the Laplacian, and noise $\xi = 0$ is to view the problem as a linear program ([3]), as in equation 10

$$\min \mathbf{c}^T|\mathbf{s}| \quad \text{subject to} \quad \mathbf{A}\mathbf{s} = \mathbf{x} \quad (10)$$

Letting $\mathbf{c} = (1,\ldots,1)$, the objective function of the linear program becomes $\mathbf{c}^T|\mathbf{s}| = \sum_m |s_m|$, equivalent to maximizing the log posterior under a Laplacian prior. Reformulating the problem as a standard linear program with only positive coefficients, it can be solved quickly and efficiently using interior point linear programming methods (such as those in MATLAB's optimization toolbox) and quadratic programming methods. A similar approach is used in [7].

5. SIMULATION RESULTS

To test the system, three underdetermined scenarios were used; the first, the extraction of four speakers from three mixtures, then the extraction of three speakers from two mixtures and finally two speakers from a single mixture. The mixtures consisted of two different male speakers, one female speaker and a piece of background music all sampled at 8 kHz. The signal dictionaries had been learned from each source type as described above. The matrix $\mathbf{A}$ was pre-determined but could be varied with minimal change in results. To determine the quality of separation, the signal-to-noise ratio was used. Comparison is made between the full dictionary, reduced dictionary and the learned basis of ([3]) using signal to noise ratio. For visual analysis, close-up waveforms showing from a single source. Thirty samples are shown for the best trade off between signal length and clarity of separate method waveforms.

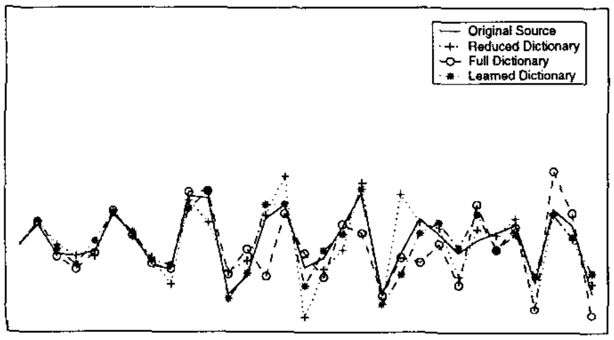

Figure 2: Separation of four sources from three mixtures. Close-up of source 1 over 30 samples together with separated sources.

The first scenario is the best determined of the three and we can see in Figure 2 it yields the best results. For the fully determined dictionary, the SNR ranges from 28dB to 30.2dB, falling to 26.5dB to 28dB for the reduced dictionary. The learned dictionary is a slight improvement on the full dictionary with results ranging from 26.8dB to 28.0dB. This demonstrates that the fixed dictionary can obtain results comparable to a learned basis.

Figure 3 shows a close-up of one of the the separated waveforms resulting from the separation of three sources from two mixtures. The SNR in this case ranges from 15.1dB-19.5dB for the sources using the reduced dictionary, whereas the learned dictionary shows SNR levels of 17.0dB to 21.1dB showing that the reduced representation is within 2dB of the learned one.

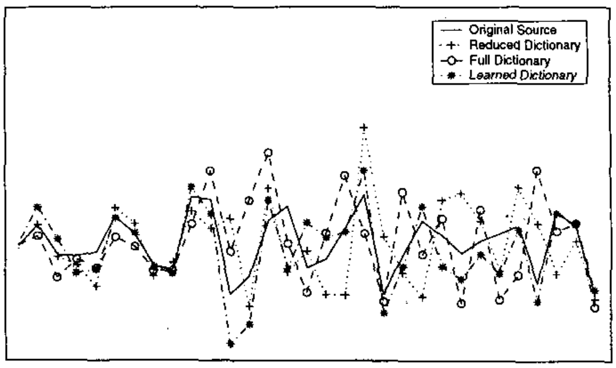

Figure 3: Separation of three sources from two mixtures. Close-up of source 1 over 30 samples together with separated sources.

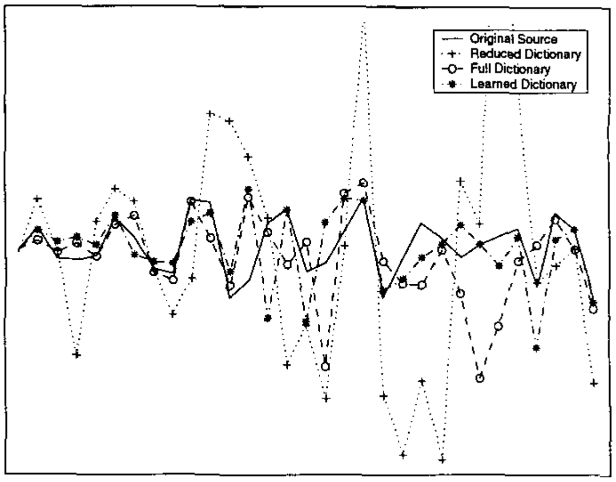

Figure 4: Separation of two sources from one mixture. Close-up of source 1 over 30 samples together with separated sources.

The third scenario results are shown in Figure 4. The results are much poorer in this instance for all three methods. The learned basis shows the SNR is around 13.1dB to 16.2dB for the extracted signals. The reduced dictionary produces lower results of 11.5dB and 14.6dB but we can see that the results are now within 1.6dB of the learned method.

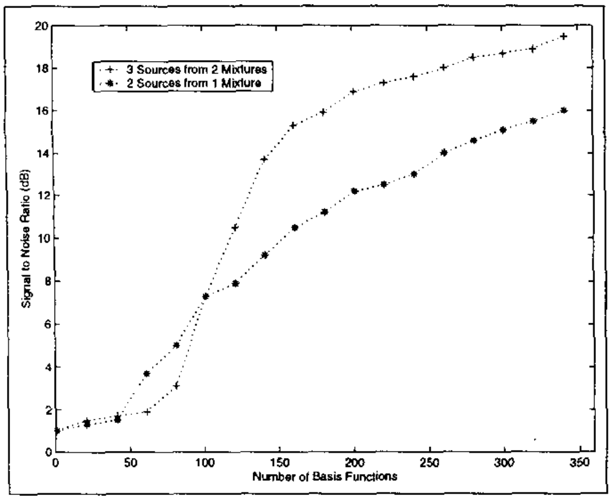

Figure 5: Results of experiments with source separation and size of dictionary. The number of basis functions used is plotted against the average SNR of the separated sources.

The graph in Figure 5 shows how reducing the size of the basis functions library effects separation SNR. We can see that using 300 basis functions rather than 150 only yields a 2.9dB improvement. Conversely, the effect of using 150 basis functions rather than 100 is an almost doubling of SNR demonstrating that this would be a good compromise library size.

In terms of perception, the recovered sources clearly have the source in question at the forefront although there are differing amounts of background noise. The first two scenarios have a very small amount of perceivable interference and the speaker dominates. In the third scenario, the background noise is much higher. Despite this, it is easy to distinguish the sentence from the other speakers and a definite degree of separation has occurred in relation to the original mixtures.

6. CONCLUSIONS

We can see that in the underdetermined case for blind separation, the use of a pre-learned signal dictionary produces comparable results to those of an optimally learned basis set using only a fraction of the processing power at run-time and a small effect on performance. This makes the system more appropriate for a practical blind separation system.

This system can be extended to use any number of different types of sources beyond speech and music. The ideal would be to compile a general purpose library for each type of source and use these sets for separation of an arbitrary source. This method could also be used as part of a feedback system where the results of separation could be used to alter the choice of basis functions.

7. REFERENCES

[1] Anthony J. Bell and Terence J. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[2] Pierre Comon. Independent component analysis - a new concept? *Signal Processing*, 36(3):287–314, 1994.

[3] Michael S. Lewicki and Terence J. Sejnowski. Learning overcomplete representations. *Neural Computation*, 12(2):337–365, 2000.

[4] Te-Won Lee et al. Blind source separation of more sources than mixtures using overcomplete representations. *IEEE Transactions on Signal Processing*, 11(2):417–441, 1999.

[5] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

[6] Michael Zibulevsky and Barak A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computation*, 13(4):863–882, 2001.

[7] Pau Bofill and Michael Zibulevsky. Underdetermined blind source separation using sparse representations. *Signal Processing*, 81(11):2353–2362, 2001.

8. ACKNOWLEDGEMENTS

The authors would like to thank the support from the U.K. MOD TG01 CRP Fellowship project.